Covariance & Cross Correlation

James R. Graham

2014/11/16

Combining Two Observations

Suppose

$$z = x + y$$

$$s_z^2 = \tfrac{1}{N-1}\sum\left(z_i - \overline{z}\right)^2, \quad \overline{z} = \overline{x} + \overline{y}$$

$$= \tfrac{N}{N-1}\overline{x^2} + \tfrac{N}{N-1}\overline{y^2} + \tfrac{2}{N-1}\sum x_i y_i - \tfrac{N}{N-1}(\overline{x})^2 - \tfrac{2N}{N-1}\overline{x}\,\overline{y} - \tfrac{N}{N-1}\left(\overline{y}\right)^2$$

$$= \underbrace{\tfrac{N}{N-1}\left[\overline{x^2} - (\overline{x})^2\right]}_{s_x^2} + \underbrace{\tfrac{N}{N-1}\left[\overline{y^2} - \left(\overline{y}\right)^2\right]}_{s_y^2} + \underbrace{\tfrac{2N}{N-1}\left(\overline{xy} - \overline{x}\,\overline{y}\right)}_{2s_{xy}}$$

$$s_z^2 = s_x^2 + s_y^2 + 2s_{xy}$$

- s_{xy} is the covariance
  - Determines the degree to which x and y are correlated

Covariance

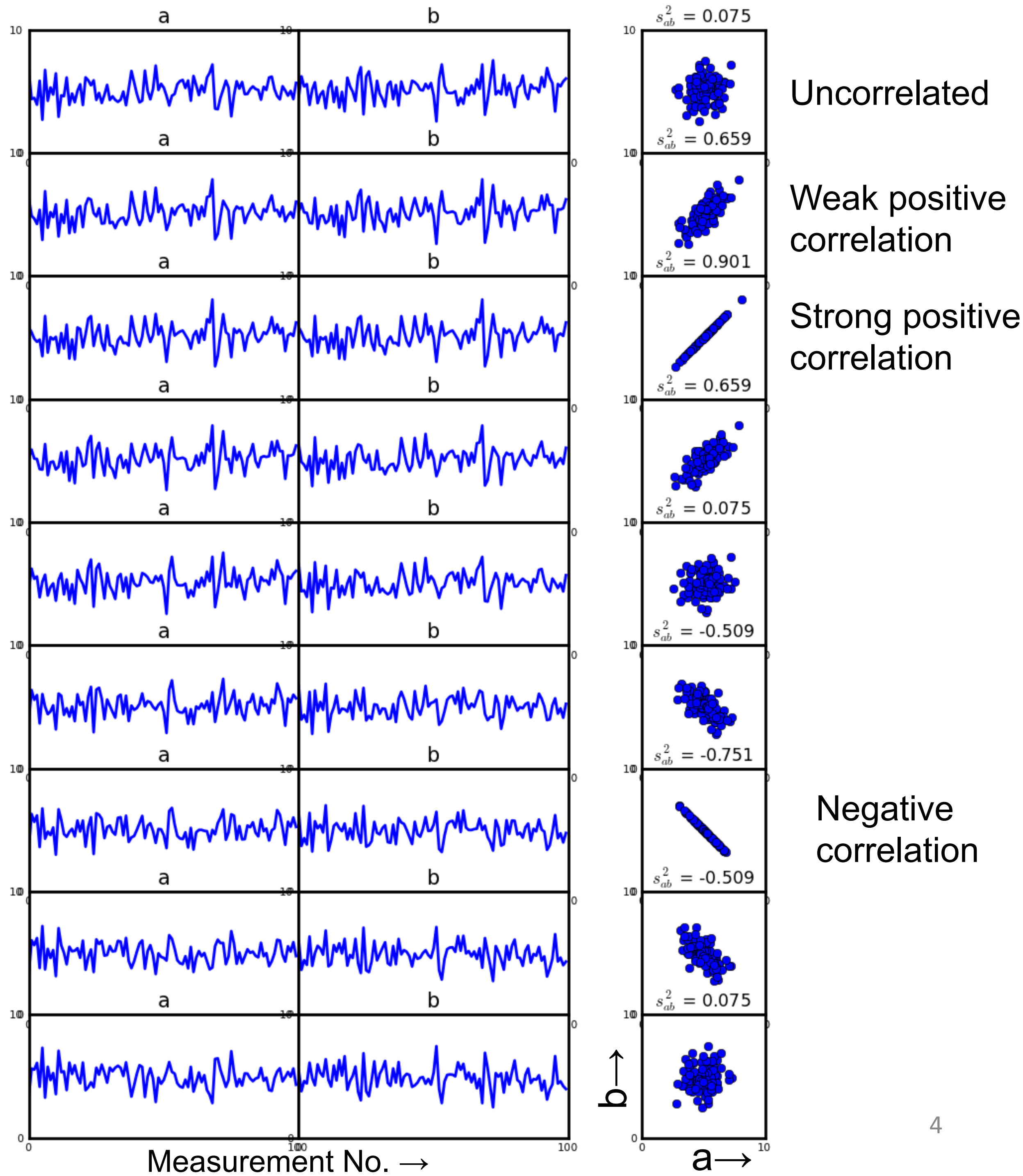
- If we have two quantities *x* and *y* the covariance is a measure of how correlated they are

$$s_{xy} = \tfrac{N}{N-1}\left[\overline{xy} - \bar{x}\,\bar{y}\right]$$

$$= \tfrac{1}{N-1}\sum_i \left(x_i y_i\right) - \tfrac{N}{N-1}\bar{x}\,\bar{y}$$

$$= \tfrac{1}{N-1}\sum_i \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$$

Correlated & Uncorrelated



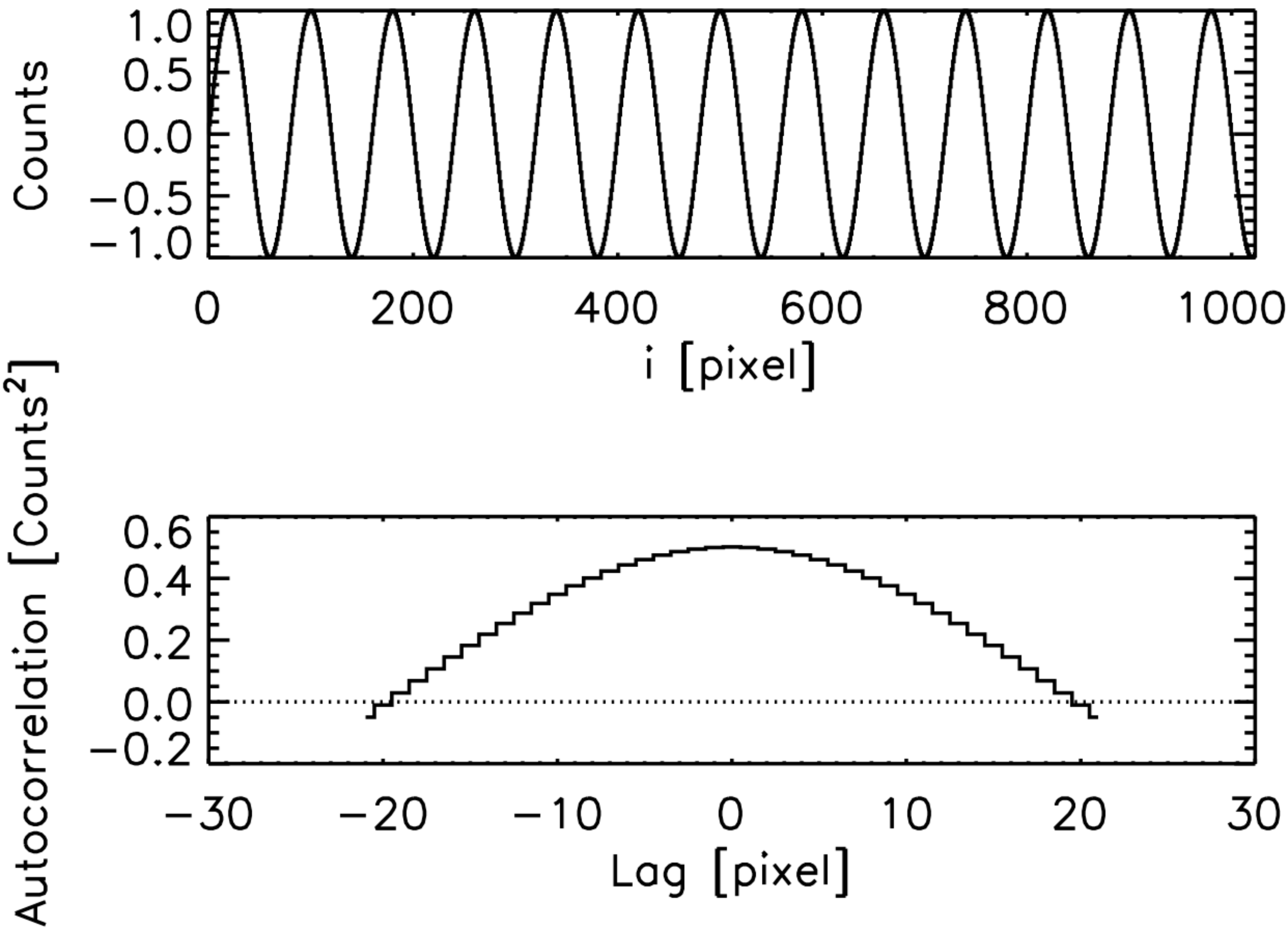
Auto-Covariance

- How can we describe the correlation of a sequence of data with itself?
  - For example with a shifted copy of itself by *j* steps

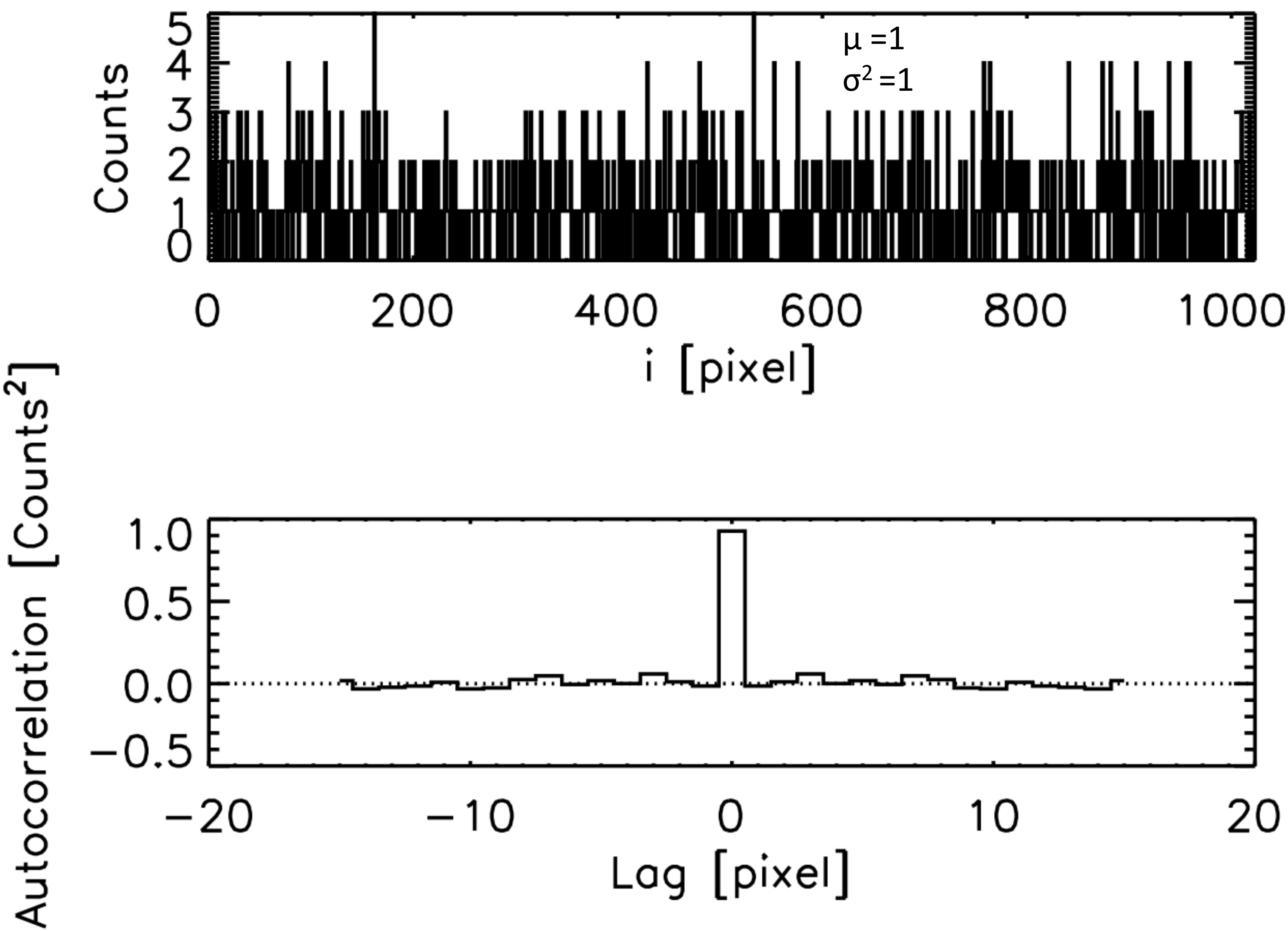
$$y_i = x_{i+j}$$

$$s_j = \tfrac{1}{N-1}\sum_i \left(x_i x_{i+j}\right) - \tfrac{N}{N-1}\left(\overline{x}\right)^2$$

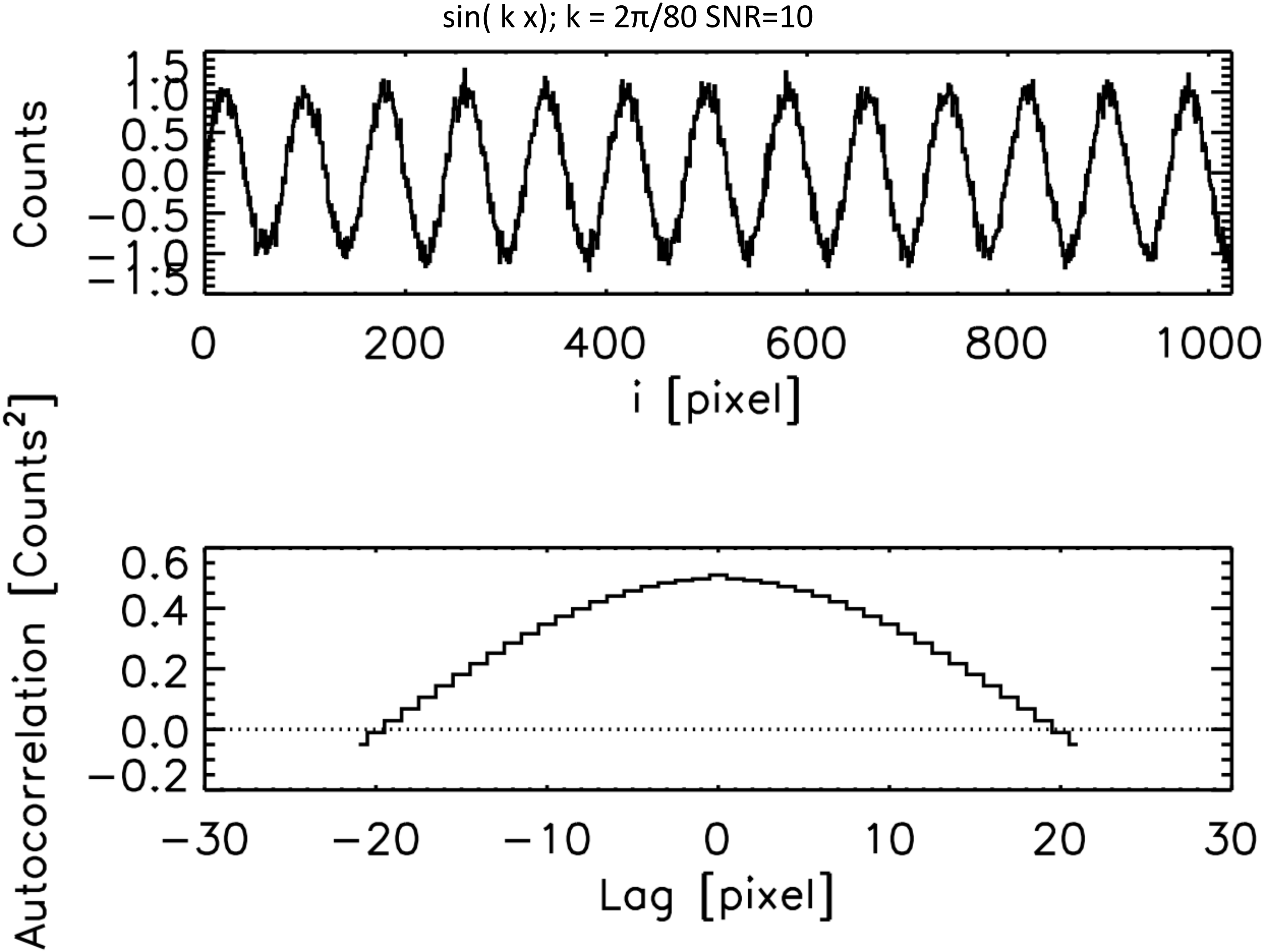
sin(k x); k = 2π/80



1.0
0.5
0.0
−0.5
−1.0
Counts
0
200
400
600
800
1000
i [pixel]
0.6
0.4
0.2
0.0
−0.2
Autocorrelation [Counts²]
−30
−20
−10
0
10
20
30
Lag [pixel]



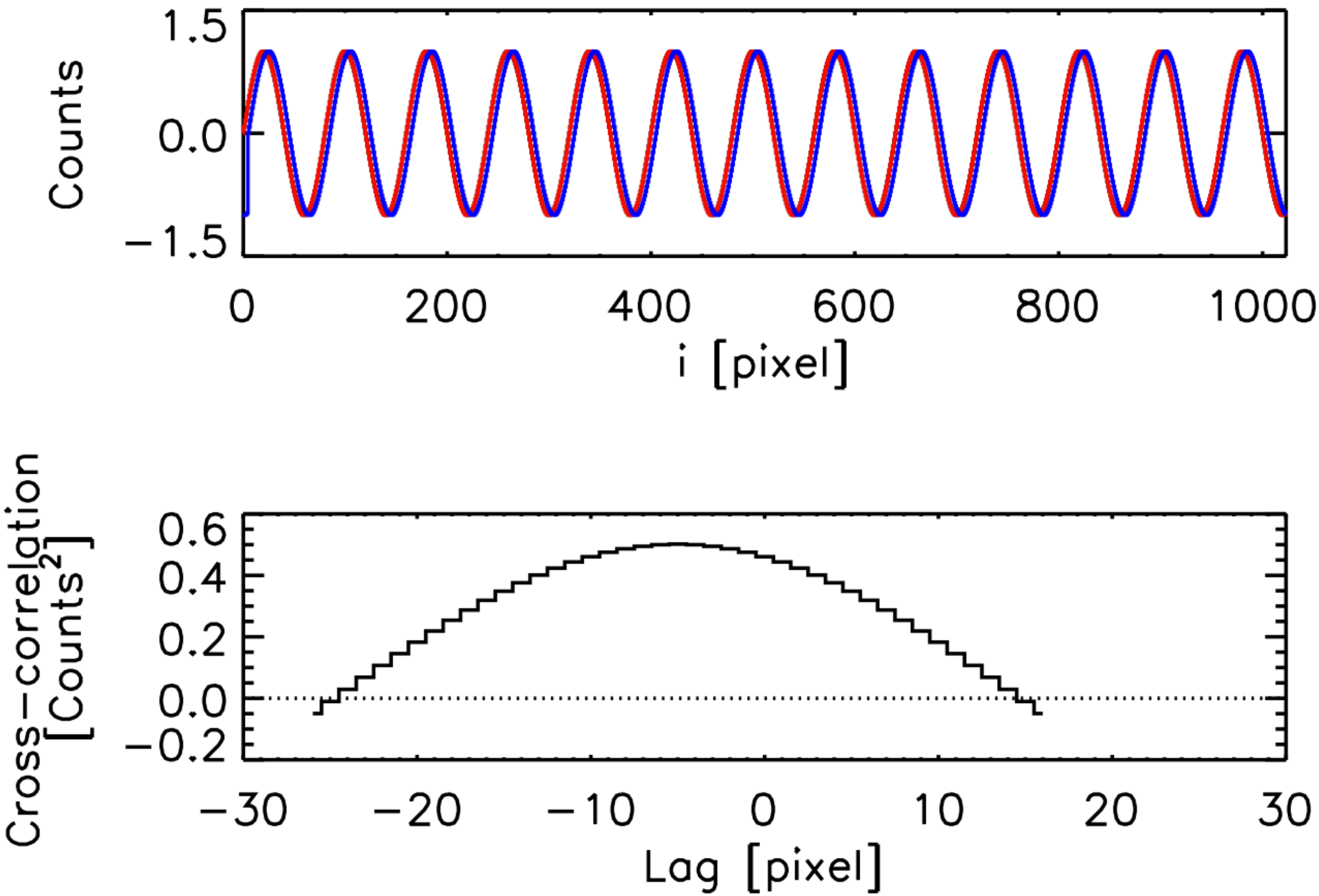
Counts
μ =1
σ² =1
i [pixel]
Autocorrelation [Counts²]
Lag [pixel]


sin(k x); k = 2π/80 SNR=10
Counts
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
200
400
600
800
1000
i [pixel]
Autocorrelation [Counts^2]
0.6
0.4
0.2
0.0
−0.2
−30
−20
−10
0
10
20
30
Lag [pixel]

Cross Correlation

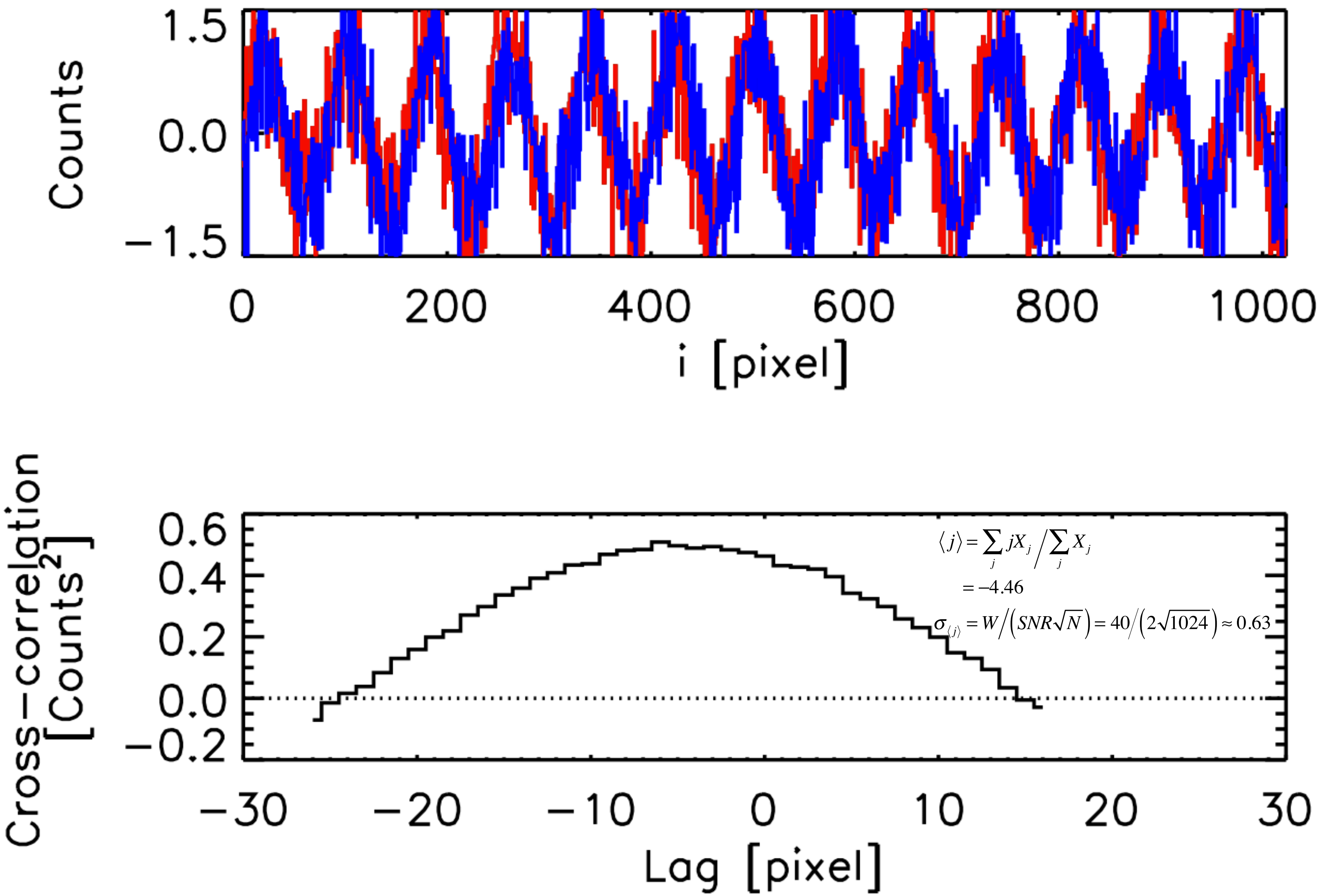
- The cross-correlation of two data sets $\{x_0, x_1, x_2 \ldots\}$ and sets $\{y_0, y_1, y_2 \ldots\}$ is defined as

$$s_j = \tfrac{1}{N-1} \sum_i \left(x_i y_{i+j} \right) - \tfrac{N}{N-1} \overline{x}\, \overline{y}$$



1.5
0.0
−1.5
Counts
0
200
400
600
800
1000
i [pixel]
0.6
0.4
0.2
0.0
−0.2
Cross−correlation [Counts^2]
−30
−20
−10
0
10
20
30
Lag [pixel]

SNR = 2



Counts
1.5
0.0
−1.5
0
200
400
600
800
1000
i [pixel]
Cross−correlation [Counts^2]
0.6
0.4
0.2
0.0
−0.2
−30
−20
−10
0
10
20
30
Lag [pixel]
$\langle j \rangle = \sum_j jX_j \Big/ \sum_j X_j$
$= -4.46$
$\sigma_{\langle j \rangle} = W \Big/ \left(SNR\sqrt{N} \right) = 40 \Big/ \left(2\sqrt{1024} \right) \approx 0.63$

Shifting an Array in Python

- Shift an array x by n elements
  - y = numpy.roll(x,n)